

AMENDMENTS TO THE CLAIMS

1. (Currently Amended) A method comprising:
receiving at a server computer system a client request from a client computer device via a network;
interpreting the client request including identifying a selection of at least one of a plurality of web interaction modes, each of the plurality of web interaction modes to perform interpretation of content being transmitted between the server computer system and the client computer device, wherein two or more of the plurality of web interaction modes are used independently or concurrently to retrieve speech processing information directly from the Internet; and
identifying a web interaction mode selected by the client computer device, and performing speech processing based on the selected web interaction mode and the retrieved speech processing information, wherein performing speech processing includes
determining an active display element that is to be focused and identifying the active display element with its associated identifier, wherein the active display element includes an element upon which a speech input received from a user is focused, the speech input is received via the client computer device,
receiving an utterance from a user, via the client computer device, once the active display element is focused, and, if the utterance matches the speech input, transmitting the identifier to the server computer system so that speech recognition is performed,

performing speech recognition based on a relationship between the active display element and one or more speech elements, wherein performing speech recognition includes retrieving a synchronization relationship between the one or more speech elements and the active display element to compose grammar of the one or more speech elements, and dynamically correcting the composed grammar of the one or more speech elements using a real-time speech recognition based on the synchronization relationship.

Claims 2-3 (Cancelled)

4. (Previously presented) The method as claimed in Claim 1 wherein the focused active element comprises a hyperlink or a field in a form.
5. (Cancelled)
6. (Previously presented) The method as claimed in Claim 1 further including: extracting speech features from a user speech input, , wherein the user speech input is contained in the client request.
7. (Cancelled)
8. (Previously presented) The method as claimed in Claim 1 further including: receiving a session message at the server computer system to initialize a connection between the server computer system and the client computer device, wherein the session message includes an internet protocol (IP) address of the client computer device, a device type of the client computer device, a voice character of a user responsible for the user speech input, a language of the user

speech input, and a default recognition accuracy requested by the client computer device.

9. (Cancelled)
10. (Previously presented) The method as claimed in Claim 1 further including:
receiving a transmission message at the server computer system to exchange
transmission parameters between the server computer system and the
client computer device.

Claims 11-13 (Cancelled)

14. (Previously presented) The method as claimed in Claim 1 further including:
receiving an exit message at the server computer system to terminate a user
session with the server computer system and the client computer device.

Claims 15-34 (Cancelled)

35. (Currently Amended) A non-transitory machine-readable medium having
instructions which when executed cause a machine to:

receive at a server computer system a client request from a client computer device
via a network;

interpret the client request including identifying a selection of at least one of a
plurality of web interaction modes, each of the plurality of web interaction
modes to perform interpretation of content being transmitted between on a
server computer system and a client computer device, wherein two or
more of the plurality of web interaction modes are used independently or
concurrently to retrieve speech processing information directly from the
Internet; and

identify a web interaction mode selected by the client computing device, and

performing speech processing based on the selected web interaction mode
and the retrieved speech processing information, wherein performing
speech processing includes
determining an active display element that is to be focused and identifying
the active display element with its associated identifier, wherein
the active display element includes an element upon which a
speech input received from a user is focused, the speech input is
received via the client computer device,
receiving an utterance from a user, via the client computer device, once
the active display element is focused, and, if the utterance matches
the speech input, transmitting the identifier to the server computer
system so that speech recognition is performed,
performing speech recognition based on a relationship between the active
display element and one or more speech elements, wherein
performing speech recognition includes retrieving a
synchronization relationship between the one or more speech
elements and the active display element to compose grammar of
the one or more speech elements, and
dynamically correcting the composed grammar using a real-time speech
recognition based on the synchronization relationship.

36. (Cancelled)

37. (Cancelled)

38. (Currently Amended) The non-transitory machine-readable medium as claimed in
Claim 35 wherein the focused active element is a hyperlink or a field in a form.

Claims 39-44 (Cancelled)

45. (Currently Amended) A system comprising:

a server computer system coupled with a client computer device, the server computer system having a storage medium and a processor coupled to the storage medium, the processor to

receive a client request from a client computer device via a network;

interpret the client request including identifying a selection of at least one of a plurality of web interaction modes, each of the plurality of web interaction modes to perform interpretation of content being transmitted between the server computer system and the client computer device, wherein two or more of the plurality of web interaction modes are used independently or concurrently to retrieve speech processing information directly from the Internet;

identify a web interaction mode selected by the client computing device, and performing speech processing based on the selected web interaction mode and the retrieved speech processing information, wherein performing speech processing includes

determining an active display element that is to be focused and identifying the active display element with its associated identifier, wherein the active display element includes an element upon which a speech input received from a user is focused, the speech input is received via the client computer device,[[;]]

receiving an utterance from a user, via the client computer device, once the active display element is focused, and, if the utterance matches

the speech input, transmitting the identifier to the server computer system so that speech recognition is performed, performing speech recognition based on a relationship between the active display element and one or more speech elements, wherein performing speech recognition includes retrieving a synchronization relationship between the one or more speech elements and the active display element to compose grammar of the one or more speech elements; and dynamically correcting the composed grammar using a real-time speech recognition based on the synchronization relationship.

46. (Previously presented) The system as claimed in Claim 45 wherein the processor is further to:

extract speech features from a user speech input, wherein the user speech input is contained in the client request.

47. (Previously presented) The system as claimed in Claim 45 wherein the processor is further to:

receive a session message at the server computer system to initialize a connection between the server computer system and the client computer device, wherein the session message includes an internet protocol (IP) address of the client computer device, a device type of the client computer device, a voice character of a user responsible for the user speech input, a language of the user speech input, and a default recognition accuracy requested by the client computer device.

48. (Previously presented) The system as claimed in Claim 45 wherein the processor is further to:

receive a transmission message at the server computer system to exchange transmission parameters between the server computer system and the client computer device.

49. (Previously presented) The method as claimed in Claim 45 wherein the processor is further to:

receive an exit message at the server computer system to terminate a user session with the server computer system and the client computer device.